

# ProtPOS: A python package for the prediction of protein preferred orientation on a surface

Jimmy C. F. Ngai<sup>1</sup>, Pui-In Mak<sup>2</sup>, and Shirley W. I. Siu<sup>\*1</sup>

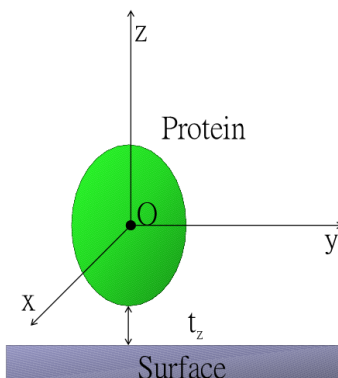
<sup>1</sup>*Computational Biology and Bioinformatics Lab, Faculty of Science and Technology, University of Macau, Macau*

<sup>2</sup>*State-Key Laboratory of Analog and Mixed-Signal VLSI and Faculty of Science and Technology, University of Macau, Macau*

March 10, 2016

## 1 The protein-surface complex system

In ProtPOS, both protein and surface molecules are treated as rigid bodies. The search of a protein preferred orientation includes changes of its position and orientation relative to the surface by translation and rotation along three coordinate axes as shown in Fig. S1. The protein positional search in the lateral direction can be limited to the dimensions of the unit cell if it is a homogeneous and periodic surface (often used in a surface model), whereas protein positional search in the surface-normal direction can be restrained to maintain within the solid-liquid interfacial region, practically by adjusting the residue minimum distance of the protein to the surface ( $t_z$ ).



**Figure S 1:** A schematic view of the protein-surface complex system. The origin of the coordinate system is defined at the protein's center of mass. Each PSO move is a combination of 3-axis translation and 3-axis rotation of the protein. To ensure a generated conformation will not result in a crash with the surface, protein translation in the Z direction will maintain a minimum distance to the surface in a given range (i.e.,  $t_z \in [\text{minz}, \text{maxz}]$ ).

## 2 Validation of ProtPOS predictions

To assess the validity of the method, ProtPOS was used to predict the preferred adsorption orientation of a 129-residue lysozyme molecule (PDB 1AKI) on a surface of perfluorodecane molecules. Due to the stochastic nature of the PSO algorithm and the non-specific adsorption behavior of proteins on a hydrophobic surface, we repeated ProtPOS for 48 times in order to

---

\*E-mail: shirleysiu@umac.mo

identify unique low-energy protein orientations. The following PSO parameters were used in ProtPOS runs:  $n=200$ ,  $c1=c2=1.193$ ,  $w=0.721$ ,  $r=10$  or reaching 200 iterations. Then, the predictions were subjected to clustering analysis based on the RMSD of residue minimum distance profiles. Finally, four clusters of different protein orientations on the surface were obtained, with 19% of predictions classified into cluster 1, 8% into cluster 2, 15% into cluster 3, and 58% into cluster 4. For each cluster, the surface contacting residues of lysozyme were identified. Here, the definition of “contacting residue” was taken from [Raffaini and Ganazzoli, 2010], that is the protein residue with a minimum distance of  $< 5 \text{ \AA}$  to the surface.

Four low-energy structures (S1-S4), one from each cluster, and three randomly selected intermediary structures generated during the PSO search (N1-N3) were subjected to MD simulations to assess the stability of the predicted orientations. The simulation conditions are described as follows: The protein-surface complex structure was fully solvated and added counterions to neutralize the system. It was then energy minimized in 1000 steps followed by 2 ns position restraint and MD simulation until 50 ns or the protein diffused away from the surface. Long-range electrostatic interactions were treated using particle-mesh Ewald (PME) whereas short-range non-bonded interactions were calculated with a cutoff of 1.2 nm. All simulations were run using TIP4P water model and OPLS-AA force field for protein and surface molecules. The integration time step was 2 fs and trajectory data was output every 1 ps. The GROMACS simulation package with version 4.5.5 was employed.

### 3 Results

Our clustering analysis yielded four major different lysozyme orientations on the hydrophobic perfluorodecane surface. The orientation with respect to the surface is referred as *side-on* or *end-on*, depending on whether the protein’s long axis is parallel or perpendicular to the surface; and the orientation of the protein’s active site is referred as *face-up* or *face-down* depending on whether the active site is facing the solvent phase or the surface. In experiments [Wertz and Santore, 2002, Hunter *et al.*, 1990] and simulation studies [Sun *et al.*, 2005, Wei *et al.*, 2012, Yu *et al.*, 2014], lysozyme was observed to preferentially adsorb in side-on orientations on hydrophobic surfaces and air/water interface. While end-on orientations were also detected [Sun *et al.*, 2005], at long times, they would be reoriented to side-on by a rollover mechanism leading to the well-known “overshooting” phenomena [Wertz and Santore, 2002]. However, experimental observations about the active site direction upon adsorption on hydrophobic surfaces are not unanimous. In hydrophobic interaction chromatography, it was found that chromatographic retention deviated only when amino acid substitutions occurred on the surface of lysozyme opposite the active cleft [Fausnaugh and Regnier, 1986], suggesting a face-up orientation. On the contrary, in the side-chain modification study with circular dichroism and bioactivity assays of adsorbed enzyme layers [Fears *et al.*, 2009], the active cleft of lysozyme in the adsorption layer could not be detected and their bioactivity was significantly reduced, suggesting a face-down orientation.

For comparison, we present the summary of four major lysozyme orientations identified by clustering analysis in Table S1. We followed the procedure suggested in [Sun *et al.*, 2005] to classify the types of orientation and active site direction (see the table caption). Remarkably, all four orientations are side-on which is in excellent agreement to experiments and previous computational studies; among which three of them have the active site facing the surface and one facing the solvent.

For each cluster, we identified the major surface contacting residues and compared them to the list of residues reported in a long atomistic MD simulation study of the same protein upon initial contact to the hydrophobic polyethylene (PE) surface [Wei *et al.*, 2011]. Matching residues are underlined in Table S1; those which were found to be in contact with the surface

**Table S 1: Lysozyme orientations, active site directions and contacting residues of the predicted orientations.** Orientation of lysozyme with respect to the surface is referred as “side-on” ( $\alpha$  close to  $0^\circ$ , where  $\alpha$  is the angle between the long axis of the protein and the surface plane) or “end-on” ( $\alpha$  close to  $90^\circ$ ); and the orientation of protein’s active site is referred as “face-up” ( $\omega$  is less than  $90^\circ$ , where  $\omega$  is the angle between the surface normal and the vector from the protein center to the active site) or face-down ( $\omega$  is greater than  $90^\circ$ ). Surface contacting residues are protein residues within 5 Å of the surface in most of the predictions. Contacting residues that were also observed in [Wei *et al.*, 2011] are indicated by underline (initial contact) and asterisk (after 100 ns MD).

Cluster	Size	$\alpha$ ( $^\circ$ )	$\omega$ ( $^\circ$ )	Orientation	Active site direction	Major contacting residues
1	9	6.23	36.97	side-on	face-up	Arg14, His15, Asn77, Thr89, Arg128*
2	4	-7.22	136.90	side-on	face-down	<u>Lys1</u> , <u>Val2</u> , <u>Lys33</u> , <u>Asn37</u> , <u>Asn39</u> , <u>Gln41</u> , <u>Ala42</u> , <u>Thr43</u> , <u>Asn44</u> , <u>Arg45</u> , <u>Thr47</u> , <u>Arg68</u>
3	7	17.85	125.48	side-on	face-down	Lys33, Phe34, Asn113, Arg114, Lys116, Gly117, Thr118, Asp119, Ala122, Trp123, Arg12
4	28	-26.79	107.34	side-on	face-down	<u>Lys1</u> , <u>Val2</u> , <u>Lys33</u> , <u>Asn37</u> , <u>Phe38</u> , <u>Asn39</u> , Gly126, Arg128*

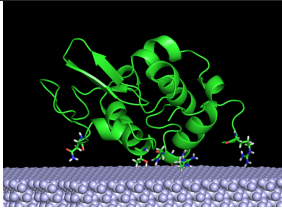
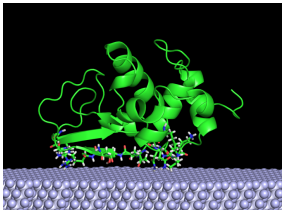
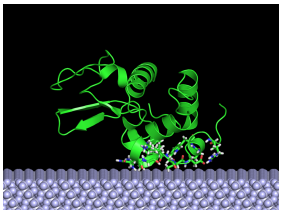
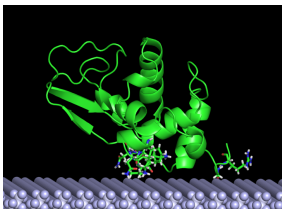
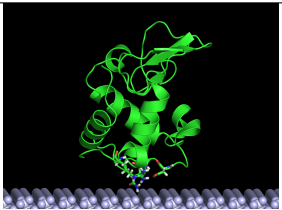
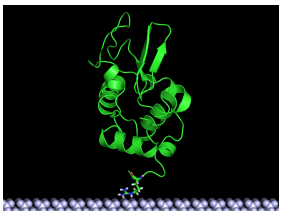
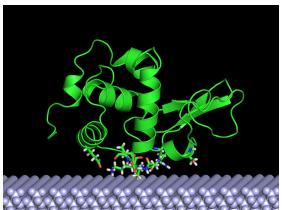
only after long simulation in [Wei *et al.*, 2011] are indicated with asterisk. It is noteworthy that all contacting residues in cluster 2 and most contacting residues in cluster 4 were observed in the MD study. In addition, the residue Arg128, which reached the PE surface only after 100 ns in the simulation, was successfully identified by ProtPOS in cluster 1 and 4.

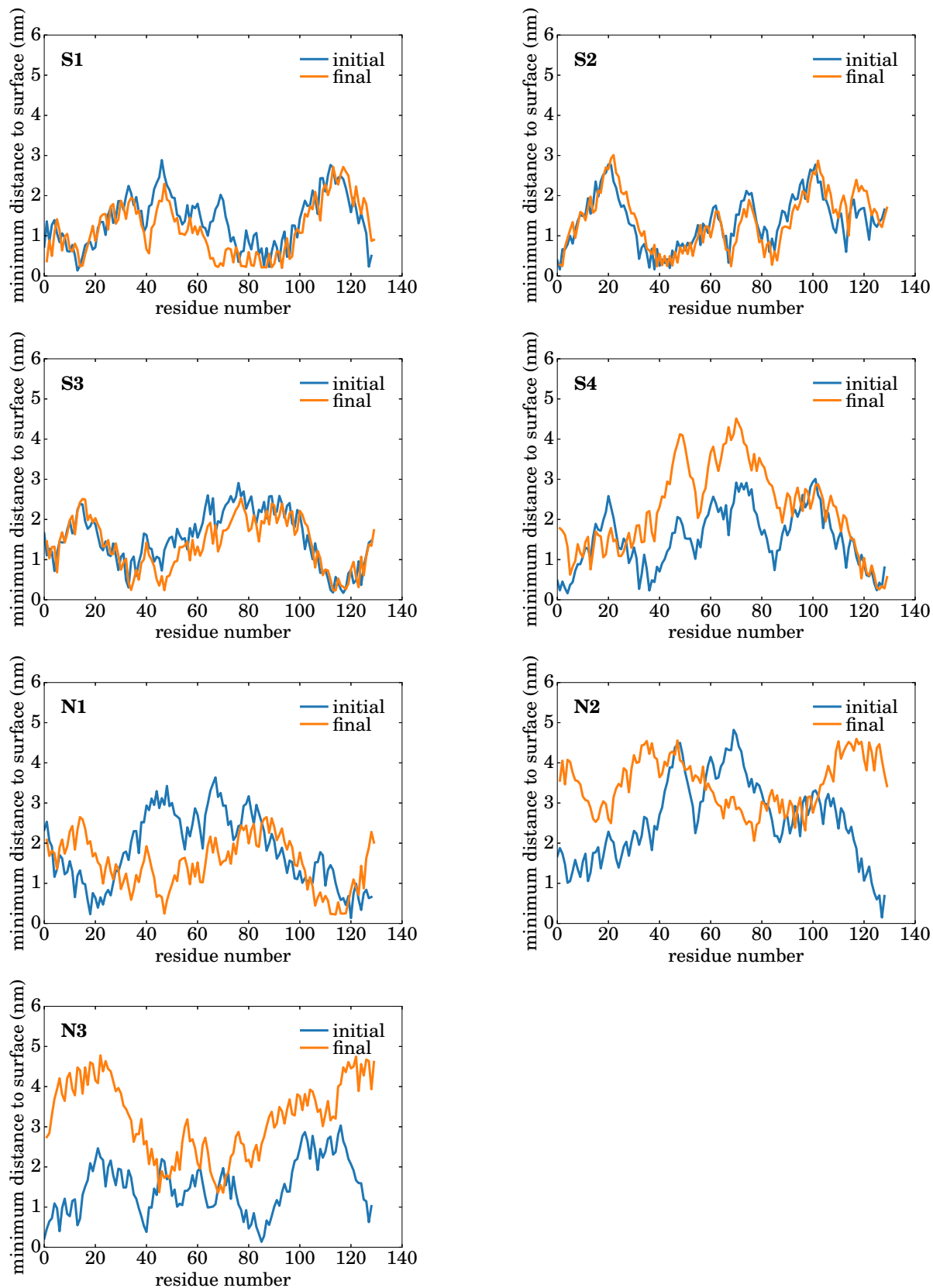
Here, stability of the predicted orientations were tested by running the fully solvated protein-surface systems to 50 ns in MD simulations. As shown in Table S2, proteins in S1-S3 maintain their predicted preferred orientation throughout the simulations. An exception is found in S4 where the protein has slightly rotated almost immediately upon the simulation began, releasing most of the surface contacting residues but not residues at the C-terminus; it maintained this new orientation until the end of the simulation. On the other hand, N systems either took longer simulation time to find out more favorable orientation to stay attached on the surface (N1) or quickly bounced-off from the surface (N2-N3).

In our model system with a homogeneous surface, the protein orientation relative to the surface can be characterized by the residue minimum distance profile. As shown in Fig. S2, new contacting residues emerged in S1-S3 systems after MD simulations due to conformational relaxation of the protein structures. Interestingly, the preferred orientation discovered in N1 simulation resembles one of our predicted orientations (S3). None of the systems reach denaturated state which is suggested to be attainable only with very long simulation time or enhanced sampling techniques such as accelerated molecular dynamics (aMD) [Wei *et al.*, 2011, Mücksch and Urbassek, 2013].

Together these results suggest that ProtPOS yields reliable structure predictions on protein preferred orientation on a surface upon initial attachment. These predictions serve as good starting structures for further simulation studies of the protein adsorption process greatly reducing the time and effort on simulating unsuccessful adsorption events.

**Table S 2: Simulations of different lysozyme orientations on the hydrophobic surface of perfluorodecane molecules:** System S1-S4 are ProtPOS predictions and N1-N3 are random selections from intermediary protein conformations during PSO sampling, sorted by their ProtPOS scores. For initial orientation images, proteins are drawn as green ribbon with contacting residues (within 5 Å to the surface) as sticks. As a summary, S systems mostly retain their predicted preferred orientations within the simulation time, except in S4 which has been rotated and re-attached to the surface with a new orientation. N systems are either resulted in re-oriented conformation or bounce-off from the surface.

System	Initial orientation	ProtPOS score (kJ mol <sup>-1</sup> )	Simulation time	Final $E_{Prot-Surf}^{Coul+LJ}$ (kJ mol <sup>-1</sup> )	Orientation and final status
S1		-1060.16	50 ns	-294.10	same and stable
S2		-959.92	50 ns	-177.45	same and stable
S3		-791.63	50 ns	-163.80	same and stable
S4		-765.23	50 ns	-86.12	changed after 3 ns then stable
N1		-241.53	50 ns	-173.98	changed after 10 ns then stable
N2		10.16	32 ns	0	bounce-off after 10 ns
N3		190.89	27 ns	0	bounce-off after 2 ns



**Figure S 2: Comparison of residue minimum distance profiles of the initial protein orientation and the final orientation after MD simulation:** For adsorbed systems S1-S3, slight adjustments of contacting residues and their distances to the surface can be observed. In S4, protein re-orientation has released most predicted protein-surface contacts but retained the contacting residues at the C-terminus. Interestingly, protein re-orientation in N1 has resulted in protein conformation resembling the prediction for S3.

## References

- [Fausnaugh and Regnier, 1986] Fausnaugh, J.L. and Regnier, F.E. (1986) Solute and mobile phase contributions to retention in hydrophobic interaction chromatography of proteins. *J Chromatogr A*, **359**, 131–146.
- [Fears *et al.*, 2009] Fears, K.P. *et al.* (2009) Probing the conformation and orientation of adsorbed enzymes using side-chain modification. *Langmuir*, **25**, 9319–9327.
- [Hunter *et al.*, 1990] Hunter, J.R. *et al.* (1990) Lysozyme adsorption at the air/water interface. *J Colloid Interface Sci*, **137**, 462–482.
- [Mücksch and Urbassek, 2013] Mücksch, C. and Urbassek, H.M. (2013) Enhancing protein adsorption simulations by using accelerated molecular dynamics. *PLoS ONE*, **8** (6), e64883.
- [Raffaini and Ganazzoli, 2010] Raffaini, G. and Ganazzoli, F. (2010) Protein adsorption on a hydrophobic surface: a molecular dynamics study of lysozyme on graphite. *Langmuir*, **26**, 5679–5689.
- [Sun *et al.*, 2005] Sun, Y. *et al.* (2005) Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation. *Langmuir*, **21**, 5616–5626.
- [Wei *et al.*, 2011] Wei, T. *et al.* (2011) Lysozyme adsorption on polyethylene surfaces: why are long simulations needed? *Langmuir*, **27** (19), 12074–12081.
- [Wei *et al.*, 2012] Wei, T. *et al.* (2012) Molecular dynamics simulation of lysozyme adsorption/desorption on hydrophobic surfaces. *J Phys Chem B*, **116**, 10189–10194.
- [Wertz and Santore, 2002] Wertz, C.F. and Santore, M.M. (2002) Adsorption and reorientation kinetics of lysozyme on hydrophobic surfaces. *Langmuir*, **18**, 1190–1199.
- [Yu *et al.*, 2014] Yu, G. *et al.* (2014) Mesoscopic coarse-grained simulations of lysozyme adsorption. *J Phys Chem B*, **118**, 4451–4460.